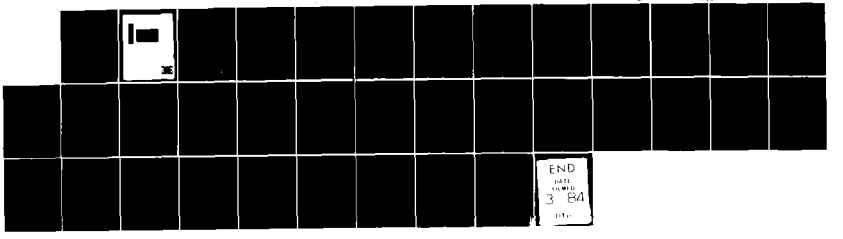
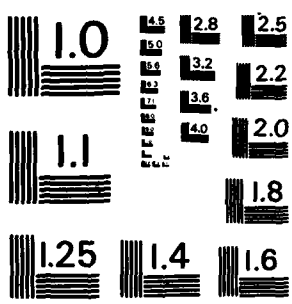
AD-A137 229 DESIGN OF EXPERIMENTS IN SIMULATION(U) DESMATICS INC 1/1
STATE COLLEGE PA D E SMITH ET AL. JAN 84 TR-113-13
N00014-79-C-0650

UNCLASSIFIED F/G 9/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DESMATICS, INC.

P. O. Box 618
State College, PA. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

DESIGN OF EXPERIMENTS IN SIMULATION

by

Dennis E. Smith
and
Carl A. Mauro

To appear as invited paper for special issue on simulation
of American Journal of Mathematical and Management Sciences

TECHNICAL REPORT NO. 113-13

January 1984

This study was supported by the Office of Naval Research
under Contract No. N00014-79-C-0650, Task No. NR 042-467

TABLE OF CONTENTS

1. INTRODUCTION

In many research areas computer simulation is a very useful and powerful technique for studying the behavior of complex real-world systems. Of course, there a number of important steps to a successful simulation study [see, for instance, Gordon (1978), Kleijnen (1974), and Law and Kelton (1982)]. Perhaps the two most critical steps are (1) model validation, and (2) the design and analysis of the simulation experiments. Careful attention to both of these steps is necessary for a meaningful and sound simulation study. For instance, if a simulation model is not sufficiently representative of the system under study (i.e., valid), the output data may be misleading and result in erroneous conclusions about the system. On the other hand, even if the model satisfactorily mimics the system, an optimal experimental design is needed in order to derive maximum benefit from the time and cost incurred.

As in any experimental investigation, a simulation study requires careful planning, data collection, output data analysis, and proper interpretation of the experimental data. In general, the problem of design is to ensure that data relevant to the proposed study is obtained in as efficient a manner as possible. We are primarily concerned in this paper with the problem of experimental design; we will assume that the simulation model is an adequate representation of the system under consideration. [General discussions of simulation validation are given by Gass (1977), Law and Kelton (1982), Naylor and Burdick (1975), Schellenberger (1974), and Van Horn (1971).] We will also assume that the computer program (code) used to execute the simulation model has been

properly debugged so that the system model functions as intended.

There are several characteristics that distinguish simulation experimentation from statistical experimentation in general. First, we have much more control over the experimental conditions than we do in the real world. This often allows us to use the simulation model to examine a number of "what if" questions about which little or no data currently exists. For example, using a simulation model of a nuclear power plant, we may wish to determine what would happen in the event of a loss-of-coolant accident. Second, we can control much of the underlying randomness in a simulation by controlling the streams of pseudorandom numbers that drive and determine the stochastic events that occur in a simulation. This capability often allows us to use variance-reduction techniques to obtain estimators having greater statistical precision. Moreover, there is generally no need for randomization of experimental conditions and run order to guard against the inadvertent introduction of systematic biases and variation. Such protection is usually provided by the random-number generators already present in the simulation model. Third, many simulation developers attempt to take into account as many detailed aspects of the system under investigation as possible. As a result, many simulation studies are characterized by the inclusion of an exceptionally large number of input variables. Finally, the problems of missing data and outliers which can handicap and reduce the effectiveness of any experimental investigation are generally of no concern in simulation studies. Outliers (i.e., discordant or contaminant observations) cannot arise because a simulation model is essentially a "closed" system; missing output data

can occur only if the time and/or funds allocated for experimentation are insufficient.

It is beyond the scope of this paper to consider all the many facets of experimental design; the current literature in this subject area is vast. Instead we choose to discuss the salient aspects of four selected topics which we feel are of particular interest and relevance in the simulation context. These are: (1) Identification of the important factors (i.e., input variables); (2) Investigation of the statistical relationship between the output and input variables; (3) Determination of the combination of factor levels for which the response (i.e., output variable) is optimized; and (4) The use of variance reduction techniques.

We address each of these topics in the ensuing sections. Throughout our discussions we assume that there is but a single response variable, and we restrict ourselves to the situation in which all the factors are quantitative.

2. FACTOR SCREENING

As noted previously, simulation models often involve a great many factors. Such models, however, because of their size and running time, can require a prohibitively large and costly experimental program to study their behavior. Therefore we may want to concentrate our analysis on the set of "most important" factors. Factor screening methods [see, for instance, Kleijnen (1975), Montgomery (1979), and Smith and Mauro (1982)] are statistical methods that attempt to identify, efficiently and economically, a set of most important factors. Once the most im-

portant factors have been identified, subsequent simulation experimentation can concentrate more intensively on these critical factors, thereby eliminating experimentation with relatively unimportant factors which can needlessly consume resources. Screening experiments, then, are not usually an end in themselves but are customarily performed as a preliminary step in the experimentation process.

In screening experiments, we want (a) to detect as many important factors as possible, (b) to declare important as few unimportant factors as possible, and (c) to expend as few runs as possible. Thus, one must generally consider both how many runs a screening strategy requires and how accurately it identifies factors. Although one may wish to obtain finer factor groupings than simply "important" or "unimportant", to effectively accomplish this would most certainly require more screening runs than are normally reasonable or affordable. In any event, the greater (lesser) the degree of importance a factor has, the larger (smaller) should be the probability of classifying that factor as important.

In screening designs, a relatively small number of factor levels is generally employed; in fact, most screening experiments are two-level experiments. There are two reasons for this. First, two levels of each factor are usually sufficient to detect which factors have major effects. Second, and more importantly, two-level designs maximize the number of factors that can be examined in a given number of runs because the number of factor level combinations is minimized when each factor has only two levels.

The full statistical model for a two-level complete factorial ex-

periment for k factors contains 2^k terms: a mean effect, k main effects, $\binom{k}{2}$ two-factor interaction effects, $\binom{k}{3}$ three-factor interaction effects, ..., a k-factor interaction effect. To estimate every effect in the full model, one must run the complete factorial experiment consisting of $N = 2^k$ runs. This many runs, however, is rarely practical in simulation experimentation; for even a moderate number of factors the implications in terms of money invested and overall run time can be quite overwhelming. However, if we can reasonably assume that certain higher order interactions are negligible, we can make a less than complete investigation by running only a fraction of the 2^k treatment combinations.

In this section, we will consider two basic situations: (1) the unsaturated/saturated case, and (2) the supersaturated case. In the first case, one can afford to invest more runs than there are factors, but still considerably less than 2^k; in the second case, the number of runs available for screening is less than or equal to the number of factors to be screened. For the remainder of this section we present and discuss those experimental plans which are particularly suited for screening in these two cases. Before proceeding with the main discussion, however, we digress momentarily to review a few fundamental terms and concepts in design theory.

2.1 Orthogonality, Confounding, and Resolution

The levels to be run in a two-level screening experiment can be conveniently displayed in a design matrix such as is given in Table 1.

Run	x_1	x_2	x_3	x_4	x_5
1	+1	+1	+1	-1	+1
2	+1	-1	-1	-1	+1
3	+1	+1	+1	-1	-1
4	-1	-1	+1	+1	-1

Table 1. A two-level design for a five-factor experiment in four runs.

We have arbitrarily coded the two levels of each factor as +1 (high) and -1 (low). In run #1, for example, all factors except x_4 are held at their high level.

When ±1 coding is used, we call two design columns orthogonal if the sum of their cross products is zero. Equivalently, two columns are orthogonal if their factor levels are balanced, i.e., are different just as often as they are the same. Orthogonality is a desirable design property because estimates of the (main) effects of orthogonal factors are independent. In other words, if one of two orthogonal factors has an effect, it cannot cause the other, perhaps erroneously, to appear to have an effect. For the design matrix in Table 1, x_1 and x_3, x_2 and x_5, and x_3 and x_4 are orthogonal.

If two design columns are not orthogonal, we call the corresponding factors confounded. When two factors are confounded, it is impossible to statistically separate their effects. In the extreme case, two factors are completely confounded if their design columns are identical or are reflections of one another. For example, in Table 1, x_1 and x_4 are completely confounded. Factors x_1 and x_2, on the other hand, although not completely confounded, are not orthogonal either. In this case, we say that they are partially confounded.

The type of confounding that a design possesses is known as its resolution. In a design of resolution R, a p-factor interaction is unconfounded with any other effect containing less than R-p factors. For example, in a resolution III design, main effects are not confounded with other main effects; in a resolution IV design, main effects are not con-

founded with other main effects or two-factor interactions. The principal implication of a resolution R design is that p-factor interactions ($p < R/2$) are estimable under the assumption that all interactions of order R-p or more are negligible.

The resolution of a design is often restricted by the number of runs that can be made. For instance, in order for all columns in a design matrix to be mutually orthogonal, the number of runs must exce- the number of factors. Consequently, we can obtain unconfounded es mates of main effects only in the unsaturated/saturated case. It f lows that in the supersaturated case, design resolution must be less .an R = III, i.e., we cannot avoid confounding main effects in some manner.

2.2 The Unsaturated/Saturated Case

We now present two types of designs that are especially useful in the unsaturated/saturated case. These are Plackett-Burman (PB) designs and resolution IV foldover designs.

2.2.1 Plackett-Burman Designs

PB designs are specially constructed two-level minimal resolution III designs for studying up to $k = 4m-1$ factors in $N = 4m$ runs. PB designs, therefore, are only available for numbers of runs that are multiples of four. Assuming that all interactions can be ignored, unbiased estimation of the k main effects is possible in a PB design. The arrangements for these designs were derived by Plackett and Burman (1946); see

also Raghavarao (1971). It can be noted that when N is a power of two, PB designs are the same as the well-known resolution III 2^{k-p} fractional factorial designs, which are discussed in detail by Box and Hunter (1961).

To analyze PB designs one can use standard analysis of variance methods and conduct formal significance testing. A useful alternative approach is to plot the estimated effects on normal probability paper. In this technique, due to Daniel (1959), negligible effects should fall approximately along a straight line, while large effects should tend to fall far from the line. The latter method of analysis is especially helpful when the design is saturated (i.e., when N = k-1 and no degrees of freedom are left to estimate experimental error) or when only a few degrees of freedom are available for estimating experimental error.

2.2.2 Resolution IV Foldover Designs

Resolution IV foldover designs are easily constructed by "folding over" a resolution III design, i.e., the design matrix $\underline{D}$ can be written as

$$\underline{D} = \begin{bmatrix} \underline{D}^* \\ -\underline{D}^* \end{bmatrix}$$

where the matrix $\underline{D}^*$ is a PB design matrix. Such designs have resolution IV and allow us to study up to k factors in N = 2k runs where N is a multiple of eight. In these designs unbiased estimates of main effects can be obtained even if two-factor interactions exist.

2.2.3 Additional Remarks

For screening in the unsaturated/saturated case, resolution III and IV designs usually suffice. A resolution IV design, of course, provides more reliable information than a resolution III design but requires twice as many runs. If the simulation user is willing to invest in more than k but less than 2k runs, he or she may wish to consider other possible main-effects designs, such as "D-optimal" designs. For construction of D-optimal designs we refer the reader to the extensive literature on these designs; see, for instance, Box and Draper (1971), Dykstra (1971), Mitchell (1974), and St. John and Draper (1975). We should remark, however, that PB designs are D-optimal for their number of runs. Another interesting design optimality criterion is that of "tr(L)-optimality" for detecting the presence of two-factor interactions. These designs are studied by Morris and Mitchell (1983).

2.3 The Supersaturated Case

The supersaturated case arises when there is a severe limitation on the number of runs available for screening. Such situations are frequently encountered in simulation studies, especially in the analysis of large-scale models. The design situation of fewer runs than factors has received relatively little attention in the statistical literature, however. In fact, the performance characteristics of the supersaturated methods presently available are largely unknown.

In the following subsections we describe four basic types of designs

that have been proposed for use in supersaturated situations. These are: random balance (RB) designs, systematic supersaturated (SS) designs, group screening (GS) designs, and RB/PB combination designs. Each of these design strategies is characterized by having an equal number of runs at the high and low levels of each factor. These designs, therefore, are of resolution II. That is, main effects are not confounded with the overall mean effect.

2.3.1 Random Balance Designs

In a two-level RB design, each column of the design matrix consists of N/2 +1's and N/2 -1's where N (an even number) denotes the total number of runs to be made. The +1's and -1's in each column are assigned randomly, making all possible combinations of N/2 +1's and N/2 -1's (there are $C^N_{N/2}$ in all) equally likely, with each column receiving an independent randomization.

The principal advantage to the RB method is its flexibility; the sample size N is fixed by the simulation analyst and can be selected independently of the number of factors, k, to be screened. A second advantage is the ease with which we can prepare RB designs regardless of the magnitudes of N and k.

There are two main disadvantages to RB sampling. The first of these is that factors are confounded to a random degree. Thus, one cannot generally control the amount of confounding or interdependence between factors. Secondly, there is no specific or unique technique for analyzing RB designs. The simplest approach is to consider each factor

separately and apply some standard analysis technique such as a normal-theory F-test. More sophisticated analysis methods include variable selection procedures such as stagewise and least-squares stepwise regression methods. For a more complete discussion of RB experimentation we refer to Anscombe (1959), Budne (1959), Satterthwaite (1959), and Youden, et. al. (1959).

2.3.2 Systematic Supersaturated Designs

Because of the random confounding that occurs in RB designs, Booth and Cox (1962) introduced two-level designs which systematically attempt to minimize confounding. Noting that not all design columns can be orthogonal when $N \leq k$, Booth and Cox constructed designs that minimize $\max_{i \neq j} |c_{ij}|$ where c_{ij} is the inner product of design columns i and j. Presumably, SS designs are the best alternative to orthogonal designs, which are, of course, impossible to construct in the supersaturated case.

Booth and Cox tabulated their designs for various values of N and $k (k \leq 36)$ and outlined, for other combinations of N and k, an iterative computer procedure for generating the required designs. They admit, however, that the cost of writing and running the program may be prohibitive if k is large. An important concern, then, with SS designs is their availability.

2.3.3 Group Screening Designs

GS designs have been studied by Li (1962), Patel (1962), and Watson

(1961). In a GS design the individual factors are partitioned into groups of suitable sizes. The groups are then tested by considering each as a single factor. Because the number of groups is generally much smaller than the original number of factors, we can usually study the group factors in a standard orthogonal design such as a PB design. Moreover, we can repeat the grouping and testing process for any number of stages. At a given stage, however, we repartition only those factors within groups determined to have significant effects in the preceding stage.

The level of a group-factor is defined by assigning the group level (e.g., +1) to all component factors. This, of course, induces complete confounding of the factors within a group, which is the basic idea. At each stage of screening we can eliminate the individual factors from those groups which appear relatively unimportant.

The main advantage of GS designs is that we can to some extent control the confounding pattern. There are two corresponding disadvantages. First, the number of runs required by a GS experiment is not fixed but is random. Second, the possibility exists that effects may cancel within a group. As a simple example, consider two factors which have effects that are negatives or near negatives of each other. If these two factors are the only important factors in a group, their effects will cancel or their combined effect may be masked by experimental error. Mauro (1983a) and Mauro and Smith (1982) have examined the cancellation problem. Their results, obtained under certain simplifying assumptions, tend to indicate that cancellation does not pose a major problem to GS.

2.3.4 RB/PB Combination Designs

An RB/PB screening plan is a two-stage strategy having an RB first-stage experiment followed by the use of a PB second-stage experiment. A factor is included in the second-stage PB design only if it is determined to have a significant effect in the first stage.

As in GS designs, a disadvantage of RB/PB designs is that the total number of runs required is random (since the number of second-stage runs is random). An advantage of these designs is that the use of a PB experiment in the second stage separates any confounding between the factors that are carried over from the RB first-stage experiment.

2.3.5 Further Discussion

Because of the lack of comparative performance data, there are currently no definitive guidelines for the selection and use of supersaturated screening methods. Nevertheless, of the supersaturated screening methods presently available, the GS method has been generally recommended. Mauro (1983b), however, has recently pointed out certain practical considerations that make group screening less attractive as a technique for factor screening.

The performance characteristics of RB and RB/PB designs have been studied by Mauro and Smith (1984). They determined that RB/PB strategies perform better than RB strategies in those situations where it is important that Type I error (i.e., the chance of classifying unimportant factors as important) be maintained at a low level. In comparing SS with

RB designs, Booth and Cox (1962) concluded that unless $k < 2N$, SS designs have little advantage over RB designs.

3. INVESTIGATING THE FUNCTIONAL RELATIONSHIP

In many cases the relationship between a simulation response y and the k factors $x_1,\ldots,x_k$ can be expressed as

$$y=g(x_1,\ldots,x_k) + \varepsilon$$

where g is an unknown function and ε denotes a random error component. We often desire to know what this relationship is. In other words, we wish to determine the functional form of $g(\underline{x})$ where $\underline{x}=(x_1,x_2,\ldots,x_k)$.

In the ensuing discussion, we assume that both y and the x_i's are not only quantitative, but also continuous. Furthermore, we assume that the error component ε is normally distributed with mean 0 and variance σ^2, where σ^2 is unknown. Thus, the expected value of any observed response y corresponding to $\underline{x}$ is:

$$E(y) = g(\underline{x}).$$

Under these assumptions response surface methodology, or RSM for short, proves valuable. RSM, which is essentially a blending of statistical experimental design and regression analysis, has its foundation in a paper by Box and Wilson (1951). The terminology "response surface" derives from the fact that the mean response lies on a surface in (k+1)-dimensional space.

In industry RSM has often been applied to two general problems

associated with response surfaces. These are:

(a) Describing the response surface in some region of interest

and (b) Determining the values of the factors which produce the optimum response.

This section addresses the former topic; the next section discusses the latter topic. A detailed description of RSM and its applications is available in Davies (1978) or Myers (1971).

Basic to RSM is the 2^{k-p} fractional factorial which is an experimental design consisting of a specific fraction ($1/2^p$) of the 2^k possible points which form a full 2^k factorial experiment. In accordance with our previous discussion, we assume that the levels of each factor in the fractional factorial are coded ±1.

Under the assumption that the function g may be expressed in a Taylor series expansion about a point $\underline{x}_0=(x_{10},\ldots,x_{k0})$, its value at a point $\underline{x}=(x_1,\ldots,x_k)$ is given by

$$g(\underline{x})=g(\underline{x}_0)+\sum_{1}^{k}(x_i-x_{i0})\left(\frac{\partial g}{\partial x_i}\right)_{\underline{x}_o} \quad +\tfrac{1}{2}\sum_{1}^{k}(x_i-x_{i0})^2\left(\frac{\partial^2 g}{\partial x_i^2}\right)_{\underline{x}_o}$$

$$+\tfrac{1}{2}\sum_{i\neq j}^{k}(x_i-x_{i0})(x_j-x_{j0})\left(\frac{\partial^2 g}{\partial x_i\,\partial x_j}\right)_{\underline{x}_o}+\ldots$$

where the notation $(\cdot)_{\underline{x}_o}$ indicates that the quantity in parentheses is to be evaluated at the point $\underline{x}_0$. It should be noted that by rearranging terms, g may be expressed as a polynomial:

$$g(x)=\beta_0+\sum_{1}^{k}\beta_i x_i+\sum_{1}^{k}\sum_{1}^{k}\beta_{ij}x_i x_j+\ldots \qquad (3.1)$$

Depending upon the region of interest to the experimenter (i.e., the simulation user), it may be possible that a first-order polynomial provides a good approximation to $g(\underline{x})$ within that region. We can use a 2^{k-p} fractional factorial of at least resolution III to fit a first-order equation. This would, of course, yield the estimate

$$\hat{g}(\underline{x})=b_0+\sum_1^k b_i x_i$$

where b_i is an estimate of β_i.

The b_i's are obtained by the method of least squares through the use of the equation

$$\underline{b}=(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y} \qquad (3.2)$$

where $\underline{b}$ is a column vector of the b_i's, $\underline{y}$ is a column vector where the $j^{\underline{th}}$ entry consists of the value of the response corresponding to the $j^{\underline{th}}$ run, and $\underline{X}$ is the matrix $\underline{X}=[\underline{1},\underline{D}]$. Here $\underline{1}$ denotes a column vector of +1's and $\underline{D}$ is the design matrix.

The estimates b_i are uncorrelated and, among all unbiased linear estimates, have minimum variance. Although other designs (e.g., the simplex designs studied by Box (1952)) also provide uncorrelated, minimum variance estimates, the 2^{k-p} fractional factorial has the added advantage of being able, by the addition of specific points, to evolve directly to a second-order design which can be used to estimate quadratic effects (the β_{ij}'s). This proves valuable if it is determined that a first-order approximation is not adequate.

Because simulation runs are usually at a premium, it is a good idea

to use the smallest possible 2^{k-p} fractional factorial of resolution III. These designs are easily obtained; rules for their generation are given in Box and Hunter (1961), for example. The number of runs, N, required by these designs is, of course, given by $N=2^{k-p}$ where p is the maximum integer selected such that $2^{k-p} > k$.

As a check on how well a first-order approximation fits the true response surface, a lack-of-fit test may be conducted. In order to test lack of fit, the center point of the fractional factorial should be run in addition to the N points in the 2^{k-p} fractional factorial. Moreover, to obtain degrees of freedom for testing lack of fit, the center point should be replicated, i.e., run a number of times. If this point is replicated m times, then there will be m-1 degrees of freedom for the appropriate error term for testing lack-of-fit.

It can easily be shown that for a 2^{k-p} fractional factorial augmented with m runs at the center point, the estimated coefficients b_0, $b_1,\ldots,b_k$ are given by

$$b_o = (\sum_1^N y_j + \sum_1^m y_{0,r})/(N+m)$$

$$\text{and} \quad b_i = \sum_1^N y_j x_{ji}/N \qquad (i=1,\ldots,k)$$

where y_j denotes the observed response for the $j^{\underline{th}}$ run in the fractional factorial

$y_{0,r}$ denotes the observed response for the $r^{\underline{th}}$ run at the center point, and

x_{ji} denotes the $(j,i)^{\underline{th}}$ entry (either a +1 or a -1) in the design matrix.

Lack of fit can be obtained and tested from an analysis of variance decomposition of the overall variation in the observed simulation runs corresponding to the fractional factorial points and to the center points. The replicated runs at the center point provide the "pure" error sum of squares given by

$$SS_e = \sum_{r=1}^{m} [y_{0,r} - (\sum_{r=1}^{m} y_{0,r})/m]/(m-1).$$

The complete partition of the total sum of squares and the N+m degrees of freedom is given in Figure 1.

It can be shown that the sum of squares due to b_i, denoted by SS_{b_i}, is given by

$$SS_{b_i} = \begin{cases} (N+m)b_0^2 & i=0 \\ Nb_i^2 & i=1,\ldots,k. \end{cases}$$

The "pure" quadratic sum of squares SS_q (resulting from contributions of terms of the form x_i^2) is given by

$$SS_q = mN(\bar{y}_F - \bar{y}_C)^2/(N+m),$$

where $\bar{y}_F$ denotes the average response over the N points in the fractional factorial and $\bar{y}_C$ denotes the average response over the m runs at the center point. We see from Figure 1 that the cross-products sum of squares SS_c, arising from the terms of the type β_{ij} $(i \neq j)$, is not available when $k=N-1$. If, however, $k<N-1$, this term may be obtained most easily by subtracting all other sums of squares from the total sum of squares, SS_t, where

Source	Degrees of Freedom	Sum of Squares	Mean Square
b_0	1	SS_{b_0}	$MS_{b_0}=SS_{b_0}$
b_1	1	SS_{b_1}	$MS_{b_1}=SS_{b_1}$
.	.	.	.
.	.	.	.
.	.	.	.
b_k	1	SS_{b_k}	$MS_{b_k}=SS_{b_k}$
Lack of Fit			
Pure Quadratic	1	SS_q	$MS_q = SS_q$
Cross Products	N-k-1	SS_c	$MS_c=SS_c/(N-k-1)$
Pure Error	m-1	SS_e	$MS_e=SS_e/(m-1)$
Total	N+m	SS_t	

Figure 1. Partition of the Sums of Squares and the Degrees of Freedom in the Fractional Factorial and Center Points

$$SS_t = \sum_{1}^{N} y_j^2 + \sum_{1}^{m} y_{0,r}^2.$$

Under the hypothesis that a first-order fit is adequate, each of the two lack-of-fit terms SS_q and SS_c should measure only random error. Therefore, the corresponding mean squares $MS_q = SS_q$ and $MS_c = SS_c/(N-k-1)$ should be approximately the same size as the pure error mean square, $MS_e = SS_e/(m-1)$.

Lack of fit may be judged by the appropriate F-tests involving the ratios MS_q/MS_e and MS_c/MS_e. These ratios may be compared with the upper α points of F-distributions with (1,m-1) and (N-k-1, m-1) degrees of freedom, respectively. The significance level α is, of course, selected by the experimenter, *although* α=.05 *is an old standby.*

If no lack of fit is indicated by either F-test, the fitted functional equation may be used, within the factorial region, as a description of the unknown function $g(\underline{x})$. A significant lack of fit, however, indicates that the first-order model does not provide an adequate explanation of the observed data. In this situation our next step would be to take curvature of the response surface into account by fitting a second-order model of the form:

$$y = \beta_0 + \sum_{1}^{k} \beta_i x_i + \sum_{1}^{k} \sum_{1}^{k} \beta_{ij} x_i x_j.$$

This may be accomplished by adding the 2k axial points $(\pm\gamma,0,\ldots,0)$, $(0,\pm\gamma,\ldots,0),\ldots,(0,0,\ldots,\pm\gamma)$ to the existing fractional factorial points and center points, in order to complete what is known as a cen-

tral composite design (CCD). This design has a number of excellent properties. For a more detailed discussion of the CCD, including how to choose the value of γ, see Myers (1971).

As an aside, it should be noted that the decision to add axial points is not made until after the data resulting from the fractional factorial and center points is analyzed. In many experimental situations this would dictate the necessity for statistical blocking because the two sets of observations are not made under homogeneous conditions. Fortunately, in simulation experiments we need not worry about this, because the underlying conditions (except, of course, for any generated random numbers) will not change.

4. OPTIMUM-SEEKING

Often the goal of simulation experimentation is not to describe the response surface in a given region, but instead to obtain an optimum response. In other words, the objective is to determine the values of $\underline{x} = (x_1,\ldots,x_k)$ that maximize (or minimize) the unknown function $g(\underline{x})$.

In a sense, this type of problem-solving situation is similar to an optimization problem to be solved by mathematical programming techniques. The major difference is that no explicit objective function is stated and, in fact, exists only implicitly in the multitude of computer instructions in the programs comprising the simulation. Thus, the task of finding the best solution cannot rely on those analytical methods which are applicable when an explicit objective function exists.

Instead, a search of the relevant factor space must be made.

In many cases the search for the best simulation response is conducted by an analyst who estimates factor values which he/she believes correspond to a reasonably good solution. The analyst then uses these values as input to the simulation and observes the corresponding response. He/she may then postulate new factor values and repeat the process a number of times. Unfortunately, the analyst's search has a tendency to turn into a trial-and-error process involving a large amount of analyst effort and computer time.

As an alternative, a search algorithm may be used for exploring the factor space. Smith (1973) examined seven search algorithms and concluded that an RSM-based tended to be the best choice. However, it is not without drawbacks. For example, an RSM-based search may yield a local optimum rather than a global optimum if local optima exist.

Optimum-seeking via RSM may involve up to four phases, which are:

(1) First-order design phase

(2) Steepest ascent phase

(3) Second-order design phase

(4) Ridge analysis phase.

In the first-order design phase we must select the initial interval of values to be considered for each factor. Estimates of the first-order effects within the initial region defined by the specified intervals may be obtained from a 2^{k-p} fractional factorial. Assuming there is no lack-of-fit, the estimated coefficients $(b_1,\ldots,b_k)$ indicate the direction in which maximum improvement in the response is predicted. This

direction is known as "the path of steepest ascent."

Simulation runs corresponding to steps out on this path are then made. These runs should be made cautiously since the prediction becomes less reliable as the distance from the initial region increases. (Selection of appropriate step size is more an art than a science. See Davies (1978) for an example.) When the observed responses worsen, the process outlined in the previous paragraph is repeated unless lack-of-fit for the first-order model is noted. In that event, the existing fractional factorial is augmented by axial points to form a CCD. In this second-order design phase the resulting estimated second-order equation may then be used to predict the factor values which yield the best response. If these values fall within the experimental region, a corresponding simulation run should be made. Otherwise, the best direction in which to proceed should be determined, with simulation runs then conducted in that direction. This involves the ridge analysis phase. Ridge analysis [Draper (1963)] is the analogue of the steepest ascent procedure used with the fitted first-order equation.

5. APPLICATION OF RSM

In the two previous sections we have only briefly outlined how RSM may be used in the simulation situation. The best bet for the simulation user who wishes an adequate background in RSM for use either in investigating the functional relationship in a given region (Section 3) or in optimum-seeking (Section 4) would be to read the statistical, rather than the simulation, literature. A thorough study of the per-

tinent sections of Cochran and Cox (1957), Davies (1978), and Myers (1971) should provide the information necessary for applying RSM techniques. Needless to say, this implies a large investment of time and effort, an investment which most people cannot afford.

There is, however, an excellent alternative. That alternative is to consult a statistician who is versed in the practical and theoretical aspects of experimental design. As an aside, it should be noted that because of the independence of RSM from the simulation itself, it is feasible to automate RSM application to a large degree. In fact, Smith (1976) has developed a modular computer program, based on RSM, for optimum-seeking in the simulation situation. This FORTRAN program, which may be used for constrained as well as unconstrained optimum-seeking, is designed to function as an executive program which may be interfaced with an existing FORTRAN-based simulation. Application of the automated RSM program requires only minor modification of any simulation with which it is to be used. Although not a panacea, this program might prove useful.

6. VARIANCE REDUCTION TECHNIQUES

In Section 1 we mentioned that simulation users can, to a certain degree, control the random number streams used in simulation experiments. The basic idea of variance reduction techniques (VRTs) is to exploit this control in order to increase the precision of the simulation results. The following two simple examples are often used to

illustrate the potentially beneficial effects of such techniques.

Example 1: Let X_1 and X_2 denote the outputs of two different system variants in the same simulation model. The statistic $W=X_1-X_2$ is an unbiased estimator of the difference between the two mean responses and has variance given by $VAR(W) = VAR(X_1)+VAR(X_2)-2COV(X_1,X_2)$. If we were to use independent streams of random numbers in the two different simulations, we would expect $COV(X_1,X_2)=0$. If, however, we were deliberately to use the same stream of random numbers in the two situations, we would expect $COV(X_1,X_2)>0$. Thus, W would have a smaller variance than would occur with independent streams.

Example 2: Let Y_1 and Y_2 denote two outputs of the same system variant in a simulation model. The statistic $Z=(Y_1+Y_2)/2$ is an unbiased estimator of the common response mean and has varaiance given by $VAR(Z)=VAR(Y_1)/4+VAR(Y_2)/4+COV(Y_1,Y_2)/2$. If we were deliberately to use random input streams that were negatively correlated, we would expect $COV(Y_1,Y_2)<0$. Here, Z would have a smaller variance than would occur with independent streams.

The two variance reduction strategies illustrated in Examples 1 and 2 are known as common random numbers and antithetic variates, respectively. These methods are the two simplest, most straightforward, and most widely applied VRTs. For an excellent discussion of these two techniques we refer the reader to Schruben (1979) and Schruben and Margolin (1978). As noted by Schruben (1979), variances are not reduced uniformly by the use of these techniques but are merely shifted from more important estimators to less important estimators. For instance, in Example 1, although $VAR(X_1-X_2)$ is decreased, $VAR(X_1+X_2)$ is increased by a corresponding amount.

In addition to common random numbers and antithetic variates, other principal VRTs include importance sampling, conditional expectations, stratified sampling, selective sampling, and control variates (or regression sampling). Detailed discussions of these techniques are presented in Kleijnen (1974) and Law and Kelton (1982). See also Wilson (1983). These techniques vary in their complexity and applicability. In general, the use of these VRTs involves replacing or modifying the original sampling procedure, or using the same sampling process but employing a more sophisticated estimator.

It has been demonstrated in the literature that VRTs (in particular, common random numbers and antithetic variates) can, when appropriately applied, significantly increase the statistical efficiency of the simulation results. For example, with a judicious selection of random number streams the variances of an analyst-specified subset of b_i's, the estimates of the β_i's obtained by least squares via equation (3.2), can show a marked decrease compared to variances resulting from using independent input streams.

The use of VRTs, however, is not without its drawbacks. First, it is not always clear if the use of a VRT will result in a variance reduction; in fact, a variance augmentation may actually result. (See, for instance, Kleijnen (1974) and Ramsay and Wright (1979).) Second, analysis of the output data is generally complicated by the use of these techniques. Third, VRTs often result in increased computing costs and analyst effort, which may offset any potential gains in efficiency.

In summary, we find ourselves somewhat ambivalent about the practical worth of VRTs in simulation. Nonetheless, we would not discourage

a potential user of these techniques. We would, however, emphasize that he/she should not only be aware of their potential advantages and disadvantages, but also be intimately acquainted with both the simulation and the VRTs under consideration.

7. REFERENCES

Anscombe, F. J. (1959). "Quick Analysis Methods for Random Balance Screening Experiments." Technometrics, 1, 195-209.

Booth, K. H. V. and Cox, D. R. (1962). "Some Systematic Supersaturated Designs." Technometrics, 4, 489-495.

Box, G. E. P. (1952). "Multi-Factor Designs of First Order." Biometrika, 39, 49-57.

Box, G. E. P. and Hunter, J. S. (1961). "The 2^{k-p} Fractional Factorial Designs, Part I and Part II." Technometrics, 3, 311-351 and 449-458.

Box, G. E. P. and Wilson, K. B. (1951). "On the Experimental Attainment of Optimum Conditions." Journal of the Royal Statistical Society, (Series B), 13, 1-38.

Box, M. J. and Draper, N. R. (1971). "Factorial Designs, The $|X'X|$ Criterion, and Some Related Matters." Technometrics, 13, 731-742.

Budne, T. A. (1959). "The Application of Random Balance Designs." Technometrics, 1, 139-155.

Cochran, W. G. and Cox, G. M. (1957). Experimental Designs (Second Edition). John Wiley and Sons, Inc., New York.

Daniel, C. (1959). "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments." Technometrics, 1, 311-341.

Davies, O. L., ed. (1978). The Design and Analysis of Industrial Experiments (Second Edition). Longman, Inc., New York.

Draper, N. R. (1963). "Ridge Analysis of Response Surfaces." Technometrics, 5, 469-479.

Dykstra, O. (1971). "The Augmentation of Experimental Data to Maximize $|X'X|$." Technometrics, 13, 682-688.

Gass, S. I. (1977). "Evaluation of Complex Models." Computers and Operations Research, 4, 25-37.

Gordon, G. (1978). System Simulation (Second Edition). Prentice-Hall, Englewood Cliffs, New Jersey.

Kleijnen, J. P. C. (1974). Statistical Techniques in Simulation (In Two Parts). Marcel Dekker, Inc., New York.

Kleijnen, J. P. C. (1975). "Screening Designs for Poly-Factor Experimentation." Technometrics, 17, 487-493.

Law, A. V. and Kelton, W. D. (1982). Simulation Modeling and Analysis. McGraw-Hill, New York.

Li, C. H. (1962). "A Sequential Method for Screening Experimental Variables." Journal of the American Statistical Association, 57, 455-477.

Mauro, C. A. (1983a). "On the Performance of Two-Stage Group Screening." Desmatics, Inc. Technical Report No. 113-9.

Mauro, C. A. (1983b). "A Comparison of Random Balance and Two-Stage Group Screening Designs - Part I." Desmatics, Inc. Technical Report No. 113-12.

Mauro, C. A. and Smith, D. E. (1982). "The Performance of Two-Stage Group Screening in Factor Screening Experiments." Technometrics, 24, 325-330.

Mauro, C. A. and Smith, D. E. (1984). "Factor Screening in Simulation: Evaluation of Two Strategies Based on Random Balance Sampling." Management Science, 30, 209-221.

Mitchell, T. J. (1974). "An Algorithm for the Construction of 'D-Optimal' Experimental Designs." Technometrics, 16, 203-210.

Montgomery, D. C. (1979). "Methods for Factor Screening in Computer Simulation Experiments." Georgia Institute of Technology technical report, Atlanta, Georgia.

Morris, M. D. and Mitchell, T. J. (1983). "Two-Level Multifactor Designs for Detecting the Presence of Interactions." Technometrics, 25, 345-355.

Myers, R. H. (1971). Response Surface Methodology. Allyn and Bacon, Inc., Boston.

Naylor, T. H. and Burdick, D. G. (1975). "The Interface Between Simulation Experiments and Real World Experiments," in: J. N. Srivastava, ed., A Survey of Statistical Designs and Linear Models. North-Holland, Amsterdam.

Patel, M. S. (1962). "Group Screening with More than Two Stages." Technometrics, 4, 209-217.

Plackett, R. L. and Burman, J. P. (1946). "The Design of Optimum Multifactor Experiments." Biometrika, 33, 305-325.

Raghavarao, D. (1971). Constructions and Combinatorial Problems in Design of Experiments. John Wiley and Sons, New York.

Ramsay, T. E. and Wright, R. D. (1979). "Variance Reduction Techniques: Some Limitations and Efficiency Limits." Modeling and Simulation, 10, 2083-2090.

Satterthwaite, F. E. (1959). "Random Balance Experimentation." Technometrics, 1, 111-137.

Schellenberger, R. E. (1974). "Criteria for Assessing Model Validity for Managerial Purposes." Decision Science, 5, 644-653.

Schruben, L. W. (1979). "Designing Correlation Induction Strategies for Simulation Experiments," in: N. R. Adam and A. Dogramaci, eds., Current Issues in Computer Simulation, Academic, New York.

Schruben, L. W. and Margolin, B. H. (1978). "Pseudorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments." Journal of the American Statistical Association, 73, 504-520.

Smith, D. E. (1973). "An Empirical Investigation of Optimum-Seeking in the Computer Simulation Situation." Operations Research, 21, 475-497.

Smith, D. E. (1976). "Automatic Optimum-Seeking Program for Digital Simulation." Simulation, 27, 27-31.

Smith, D. E. and Mauro, C. A. (1982). "Factor Screening in Computer Simulation." Simulation, 38, 49-54.

St. John, R. C. and Draper, N. R. (1975). "D-Optimality for Regression Designs: A Review." Technometrics, 17, 15-23.

Van Horn, R. L. (1971). "Validation of Simulation Results." Management Science, 17, 247-258.

Watson, G. S. (1961). "A Study of the Group Screening Method." Technometrics, 3, 371-388.

Wilson, J. R. (1983). "Variance Reduction: The Current State." Mathematics and Computers in Simulation, 25, 55-59.

Youden, W. J., Kempthorne, O., Tukey, J. W., Box, G. E. P., and Hunter, J. S. (1959). "Discussions of the Papers of Messrs, Satterthwaite and Budne." Technometrics, 1, 157-193.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 113-13	2. GOVT ACCESSION NO. AD-A137 229	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DESIGN OF EXPERIMENTS IN SIMULATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dennis E. Smith and Carl A. Mauro		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0650
9. PERFORMING ORGANIZATION NAME AND ADDRESS Desmatics, Inc. P.O. Box 618 State College, PA 16804		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-467
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE January 1984 13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this report is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Experimental Design
Computer Simulation
Factor Screening
Investigating Statistical Relationships
Optimization of Simulation Outputs
Variance Reduction Techniques

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

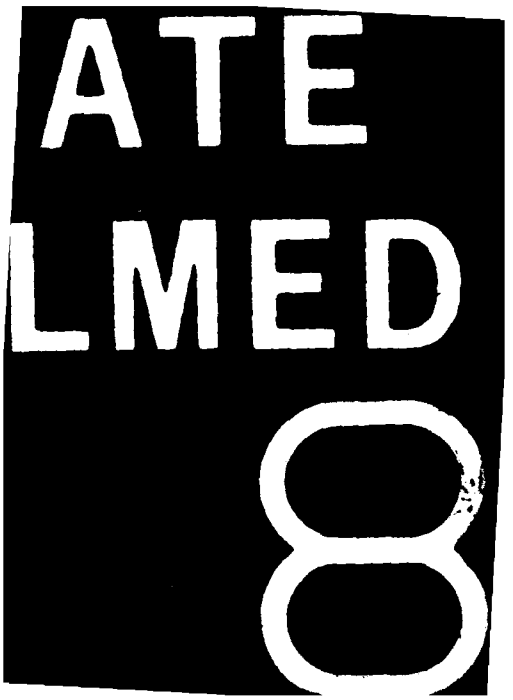
In many cases a simulation study may be viewed as an experimental situation in which a number of factors (independent variables) are to be investigated. There are, however, a number of special characteristics that distinguish simulation experiments from statistical experiments in general. In this paper we discuss three topics which we feel are of particular interest and relevance in the simulation context. These are: (1) Identification of the important factors (i.e., input variables); (2) Investigation of the statistical → cont'd

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

relationship between the output and input variables; (3) Determination of the combination of factor levels for which the response (i.e., output variable) is optimized; and (4) The use of variance reduction techniques.


ATE
LMED
8